Human KLF17 is a new member of the Sp/KLF family of transcription factors

Jane van Vliet [a], Linda A. Crofts [a], Kate G.R. Quinlan [a], Robert Czolij [a],
Andrew C. Perkins [b], Merlin Crossley [a,*]

[a] School of Molecular and Microbial Biosciences, G08, University of Sydney, Sydney, NSW 2006, Australia
[b] Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD, Australia

Abstract

The Sp/KLF transcription factors perform a variety of biological functions, but are related in that they bind GC-box and CACCC-box sequences in DNA via a highly conserved DNA-binding domain. A database homology search, using the zinc finger DNA-binding domain characteristic of the family, has identified human KLF17 as a new family member that is most closely related to KLFs 1–8 and 12. *KLF17* appears to be the human orthologue of the previously reported mouse gene, *zinc finger protein 393* (*Zfp393*), although it has diverged significantly. The DNA-binding domain is the most conserved region, suggesting that both the murine and the human forms recognize the same binding sites in DNA and may retain similar functions. We show that human KLF17 can bind G/C-rich sites via its zinc fingers and is able to activate transcription from CACCC-box elements. This is the first report of the DNA-binding characteristics and transactivation activity of human KLF17, which, together with the homology it displays to other KLF proteins, put it in the Sp/KLF family.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Krüppel-like factor; Zinc finger protein 393; ZNF393; CACCC box; GC box

There are currently 25 known members of the mammalian Sp/KLF family, which comprises Sp1–9, named for their similarity to the founding member Sp1 [1], and the Sp1-related Krüppel-like factors (KLFs) 1–16, so named for the similarity of their DNA-binding domains (DBDs) to that of *Drosophila* Krüppel [2]. The physiological roles of the Sp/KLFs are diverse [3], which is probably a reflection of the unique N-terminal activation/repression domains of each member, differences in their tissue distribution and posttranslational regulation, as well as subtle differences in DNA-binding specificities [4].

The Sp/KLFs are characterized by a highly conserved DBD at or near the C-terminus [5,6]. Their N-terminal regions, by contrast, are highly variable. The DBD consists of three C_2H_2 Krüppel-like zinc fingers that conform to the sequence found in the *Drosophila* regulatory protein Krüppel, $C-X_{2-5}-C-X_3-(F/Y)-X_5-\psi-X_2-H-X_{3-5}-H$, in which X represents any amino acid and ψ is a hydrophobic residue [7,8]. Within the Sp/KLF family, the length of each finger is invariant: the first two zinc fingers are 23 amino acids long and the third is 21 amino acids long. Four amino acids within each zinc finger, namely KS(H/L)(A/S) (finger 1), RDER (finger 2), and R(D/S)(H/R)(K/L) (finger 3), are critical for contacting DNA [9]. These DNA-contacting residues are highly conserved, suggesting that all the family members recognize sequences of the same general form, that is NNCNCNCCCN. This consensus is found in G/C-rich sites such as CACCC boxes, GC boxes, and basic transcription elements. It is thought that all Sp/KLF family members are able to bind, with varying affinities, to these sites.

The linkers separating the zinc fingers of the Sp/KLFs, which control the spacing of the fingers along the DNA, are conserved in both length and sequence, with the consensus TGE (R/K)(P/K/R)(F/Y)X [5]. It is thought that, at the start of mitosis, a kinase may inactivate all DNA-binding domains containing such linkers, thus causing a large-scale arrest in transcription [10,11].

* Corresponding author. Fax: +61 2 9351 4726.
  E-mail address: m.crossley@mmb.usyd.edu.au (M. Crossley).

In addition to binding DNA, the zinc fingers in Sp/KLFs function as a nuclear localization sequence [12–14] and a protein-interaction surface. Zinc fingers of individual family members have been shown to interact with acetylases, deacetylases, SWI/SNF chromatin-remodeling complexes, cell cycle regulatory factors, and other zinc finger proteins [15–18].

We performed a homology screen of mammalian EST databases to identify previously unrecognized Sp/KLF family members. A gene designated hypothetical protein FLJ40160 encoding a protein with the three C-terminal Krüppel-like zinc fingers characteristic of the Sp/KLF family was identified, and we now refer to it as *KLF17*. The murine homologue of *KLF17* is a previously described gene, *Zfp393* [19]. Interestingly, *KLF17* has not been well conserved between mice and humans, although it contains some conserved regions, most notably its DNA-binding domain. A phylogenetic analysis of the Sp/KLF family places *KLF17* in a branch of the KLFs. We show that KLF17 binds preferentially to the β-globin promoter CACCC box over a GC box, which is also indicative of it being a typical member of the KLF subfamily rather than one of the Sp subfamily. We show through site selection experiments that the consensus binding site for KLF17 is CCNC(C/G/a)CCC(C/g/a), which fits within the general form that is recognized by members of the Sp/KLF family. Transient transfection assays show that KLF17 is able to activate the transcription of a reporter gene from CACCC boxes.

Results and discussion

Identification of KLF17

To identify previously unreported KLF family members, the Basic Local Alignment Search Tool (BLAST) was used to search human and mouse expressed sequence tag (EST) databases (GenBank) for sequences with similarity to the DNA-binding domain of Klf3. An EST was identified that is expressed from a gene on human chromosome 1 originally annotated as hypothetical protein FLJ40160. During the course of this work it was subsequently renamed *ZNF393* (GenBank Accession No. BC049844) as part of the ongoing National Institutes of Health Mammalian Gene Collection Program effort to identify and sequence at least one full ORF-containing cDNA clone for each human and mouse gene (http://www.mgc.nci.nih. gov/ [20]) but is now referred to as *KLF17*. The 1170-nt open reading frame is predicted to encode a protein of 389 amino acids, with three C-terminal Krüppel-like zinc fingers characteristic of the Sp/KLF family, referred to as KLF17.

The human gene encoding KLF17 is shown in Fig. 1. This gene is unusual in that the termination codon lies upstream of its final exon. Stop codons are generally identified as premature when they are located more than 50–55 nt upstream of the final exon–exon junction, and the mRNA is targeted for nonsense-mediated mRNA decay (NMD) [21]. Thus almost all authentic termination codons are located in the terminal exon of mammalian genes. As the stop codon for *KLF17* is located immediately upstream of the final exon–exon junction, it is therefore not likely to be a target for NMD.

A comparison of the sequences of human *KLF17* in GenBank suggests that this gene contains two single nucleotide polymorphisms. One is a silent A→G substitution at nt 541 and the other is a G→A substitution at nt 525 that results in the amino acid substitution serine→asparagine at residue 156.

KLF17 has a processed pseudogene on human chromosome 12, denoted ψ*KLF17*, which was identified using a BLAST search of the human genome for sequences similar to human *KLF17*. Processed pseudogenes, which arise from the reverse transcription of mRNA and reintegration into the genome [22], are characterized by the absence of both 5′-promoter sequence and introns and the presence of a 3′ poly(A) tract and usually contain accumulated mutations. ψ*KLF17* shares 90% sequence identity with human *KLF17* and has a poly(A) tract, no introns, and four frameshifts and four stop codons in the potential coding sequence.

KLF17 is the human homologue of Zfp393

It has been estimated that roughly 96% of mouse genes have orthologues in humans located on syntenic intervals [23]. *KLF17* is on a region of human chromosome 1 (1p34.1) that is syntenic with a region on murine chromosome 4 (116.0–116.2 Mb). A comparison of these regions indicates that the previously reported gene *Zfp393* [19] is likely to be the murine orthologue of the human gene *KLF17*. This is supported by the observation that, when reciprocal database homology searches are performed, *Zfp393* is the murine gene with the most sequence similarity to *KLF17*, and *KLF17* is the human gene with the most sequence similarity to *Zfp393*.
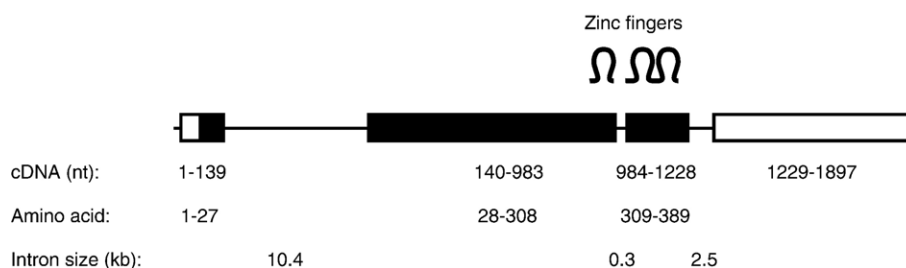

Fig. 1. The human *KLF17* gene. *KLF17* (annotated as ZNF393, GenBank Accession No. BC049844) is located on human chromosome 1p34.1. Exons are represented as boxes with the protein-coding sequences shaded black and the location of the zinc fingers shown above. The corresponding nucleotides in the cDNA and the encoded amino acid residues, along with intron sizes, are shown below. Figure is not to scale.

Table 1
In comparison with KLFs 1–16, the murine and human orthologues of KLF17 are not well conserved

Protein	Similarity of murine and human orthologues (%)	
	Non-zinc-finger region	Zinc fingers
KLF1	71.7	90.2*
KLF2	87.1	97.6
KLF3	95.8	100.0**
KLF4	89.4	100.0**
KLF5	93.0	96.4
KLF6	93.0	100.0**
KLF7	97.3	100.0**
KLF8	80.8	100.0**
KLF9	96.3	100.0**
KLF10	85.9	98.8
KLF11	98.1**	98.8
	75.7	97.5
KLF12	95.9	97.6
KLF13	92.6	100.0**
KLF14	58.3*	96.3
KLF15	84.1	100.0**
KLF16	88.5	92.6
KLF17	46.4	81.5
Average (KLFs 1–16)	87.3	98.0

The degree of conservation between human and murine orthologues of the KLFs, both within their DNA-binding domains and elsewhere, is shown. The orthologues of KLFs 1–16 with the least conservation are indicated by a single asterisk and those with the most conservation are indicated by a double asterisk. The murine orthologue of KLF14 has not been officially identified so the orthologue nominated by HomoloGene at GenBank was used. Two murine orthologues of KLF11 have been included [25].

Human KLF17 and murine Zfp393 are 54.8% similar over the entire lengths of these proteins. They exhibit higher sequence conservation in their zinc fingers (81.5% similarity)

than in the remainder of their sequence (46.4% similarity). The level of sequence similarity between human KLF17 and murine Zfp393 is low relative to the degree of conservation between orthologues in the mouse and human proteomes generally, and the KLF proteins specifically. A comparison of the mouse and human proteomes made by the International Mouse Genome Sequencing Consortium has shown that orthologues are, on average, 70% identical, and that the median identity of well-characterized nuclear-localized domains, such as zinc fingers, is 99% between mouse and human orthologues, with 70% of orthologous pairs having between 86 and 100% identity in their nuclear domains [23]. This general observation holds true for the mouse and human homologues of KLFs 1–16. These are, on average, 98.0% similar in their zinc finger domains and 87.3% similar in their non-zinc-finger regions (Table 1).

The conservation of the human and murine homologues of KLF17 also appears low compared to the degree of conservation between KLF proteins in the human and chicken. The KLFs in the chicken that have been reported to date (specifically KLFs 2, 3, 4, 5, 9, 11, 12, 13, and 15) are, on average, 70% similar over their whole sequence to the human orthologues, with their zinc fingers being 88–100% similar [24].

Fig. 2 shows the residues that have been conserved between the human and the murine forms of KLF17. In their zinc fingers, the positions of all of the zinc-chelating cysteine and histidine residues have been conserved. Moreover, all of the 12 residues at the positions believed to be the most important determinants of DNA-binding site specificity have been conserved between the two species. There are other conserved regions in the proteins that are of unknown function but may be important for structural integrity, for protein–protein interactions, or as sites of posttranslational modification.

```
Human  MYGRPQAEMECEAGELSRWQAAHQAAQDNENSAPIINMSSSSGSSGVHTSWNQGLPTIQH
Mouse  -------MECDNKE----QAMHQPPMDNKMLVPVSNVPVSSGNSGFHQP-----PATQY

Human  FPHSAEMLG PLVSVEAPGQNVNEGGPQFSMPLPERGMSYCPQATLTPSRMIYCQRMSPP
Mouse  LP---EMMR YMASAEELRCNEREWESQLIRSLPEHGVR-CPSQLAPIPFQNYCOR----

Human  QQEMTIFSGPQLMPVGEPNIPRVARPFGGNLRMPPSGLPVSASTGIPIMSHTGNPPVPYP
Mouse  ----SIGRGSHVMPVGSSGTLGVTISFSENL-MPQGGLP--SSRGVSVMAHSGAPAMPYP

Human  GLSTVPSDE------TLLGETVPSTEAQAVLPSMAQMLPPQDAHDLGMPPAESQSLLVLG
Mouse  MPPTVPATTGSLKHGILLVPGMASAGTHAVAPFMDOMLHSINPCNPEMLPARFQQLLPLD

Human  SQDSLVSQPDSQEGPFLPEQPGPAPQTVEKNSRPQEGTGRRGSSEARPYCCNYENCGKAY
Mouse  SQDSLVTESNTQEEPFVREPPTPAPEGAESPS-TSRGATRRQSPVSRPYVCTYNSCGKSY

Human  TKRSHLVSHQRKHTGERPYSCNWESCSWSFFRSDELRRHMRVHTRYRPYKCDQCSREFMR
Mouse  TKRSHLVSHQRKHTGVKPFACDWNGCTWKFFRSDELGRHKRIHTRYRPHKCDECDREFMR

Human  SDHLKQHQKTHREGPSDPQANNNNGEQDSPPAAGP
Mouse  SDHLRQHKRTHLPK-------------------
```
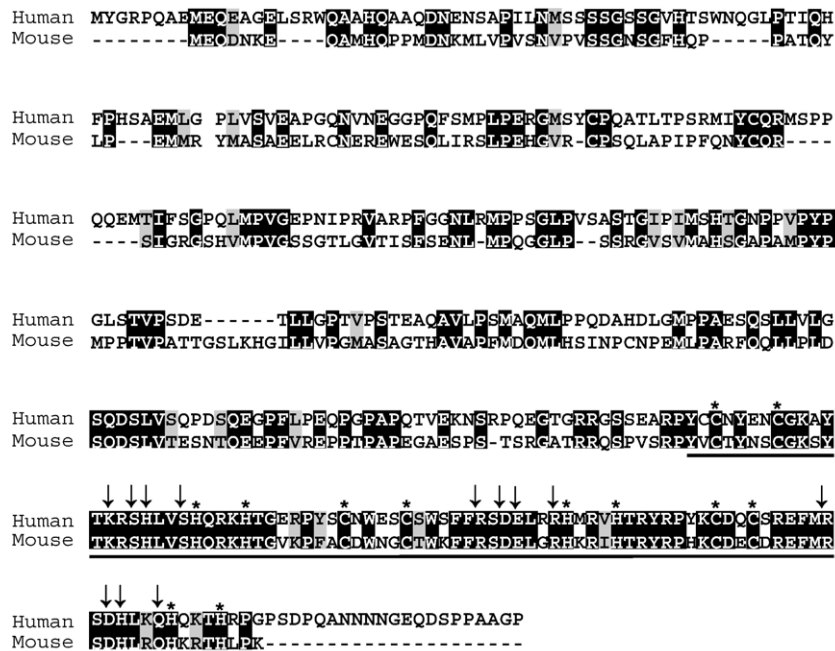
Fig. 2. Residues that are conserved between human and mouse orthologues of KLF17. Identical residues are highlighted in black and similar residues are highlighted in gray. The zinc finger domains of the human and mouse orthologues are underlined. Within the zinc finger domains, the zinc-chelating residues are marked with asterisks and the residues thought to be most important in determining DNA-binding specificity are indicated with arrows.

The relatively low level of conservation between the apparent human and murine orthologues of *KLF17* raises the possibility that there might be other homologues of *KLF17* in mice and/or humans that display more sequence conservation across the two organisms. Remarkably, a comparable example has been reported in the literature. Human *KLF11* has a homologue, *KLF11b,* on the syntenous interval in the mouse. The proteins encoded by these genes are 75% identical. There is a second homologue of *KLF11* in the mouse, presumably formed by gene duplication, which is located on a different chromosome. Of the two murine forms, this one shares the most identity with the human gene, exhibiting 98% identity at the amino acid level ([25], and refer to the sequence with GenBank Accession No. NP_848134). At present it would seem that human *KLF17* and mouse *Zfp393* are true orthologues, but the complete sequencing of the mouse and human genomes should clarify whether, as in the case of *KLF11*, there are other homologues of *KLF17* that exhibit more sequence conservation across the two species than those already identified.

Expression of KLF17

It has been reported that the expression of murine *Zfp393* is restricted to ova and sperm [19], although *Zfp393* ESTs have been derived from the early embryo, suggesting that its mRNA is also present there. The murine Zfp393 protein has not been studied but it is believed to have a role in gametogenesis or early embryogenesis, based on the distribution of its mRNA. The sources of human *KLF17* ESTs indicate that it is expressed in the testis, like the murine homologue, and also in the brain and bone. There are only a few human *KLF17* ESTs documented, suggesting that it is not highly expressed.

KLF17 is a member of the Sp/KLF family

A BLAST scan of the GenBank nr nucleotide sequence database revealed that the zinc fingers of human KLF17 most resemble, in order of decreasing similarity, those of murine Zfp393, KLFs 1–16, the Sp proteins, Wilms tumor protein 1 (WT1), Yin Yang 1 (YY1), and early growth response 1 (EGR1; Fig. 3). The zinc fingers of KLF17 are roughly 60–65% similar to those of KLFs 1–16. This is not a high level of similarity for this family; for example, the zinc fingers of KLF2 are 69–95% similar to those of KLFs 1 and 3–16. However, 11 of the 12 key DNA-binding residues in KLFs 1–16 are conserved in KLF17. The nonconserved residue is the most C-terminal of these and is Q in KLF17, not K or L, as is present in the other KLFs. Additionally, the linkers joining the zinc fingers of KLF17 are of the correct sequence and length for the Sp/KLF family.

Finally, the proteins other than the Sp/KLFs that have zinc finger domains most similar to KLF17, namely WT1, YY1, and EGR1 (Fig. 3), differ from it and the other Sp/KLFs in significant ways. WT1 has four zinc fingers, not three like the Sp/KLFs, and at least one isoform has a longer linker than that of the Sp/KLFs [26]. Additionally, 3 of the 12 critical DNA-binding residues of the Sp/KLFs are not conserved in zinc fingers 1–3 of WT1. YY1 also has four zinc fingers, with only 1 of the 12 key DNA-binding residues of the Sp/KLFs conserved in fingers 2–4 [27] and finger 2 of YY1 differs in length from the corresponding finger of the Sp/KLFs. Only 4 of the 12 critical DNA-binding residues are conserved in the zinc fingers of EGR1 [28]. Finger 3 of EGR1 does not conform to the Krüppel-like consensus and the length of finger 2 is different from that of the finger 2 of the Sp/KLFs.

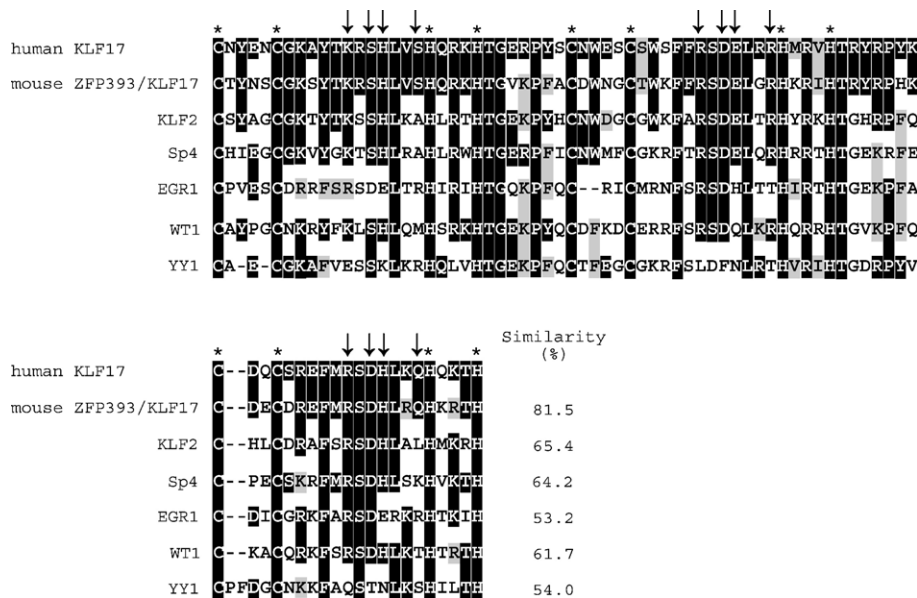It has been suggested previously that the murine protein Zfp393 is not a member of the Sp/KLF family [19]. In this

Fig. 3. The zinc fingers of KLF17 most resemble those of Sp/KLFs, Wilms tumor protein 1 (WT1), Yin Yang 1 (YY1), and early growth response 1 (EGR1). KLF2 and Sp4 are included as representatives of the Sp/KLFs. Residues identical and similar to human KLF17 are shaded black and gray, respectively. The zinc-chelating residues are marked with an asterisk and residues important for determining DNA-binding specificity are marked by arrows. WT1 and YY1 both have four zinc fingers. WT1 fingers 1–3 and YY1 fingers 2–4 are shown as these are the most similar to those of KLF17. All of the sequences, except mouse Zfp393, are human.
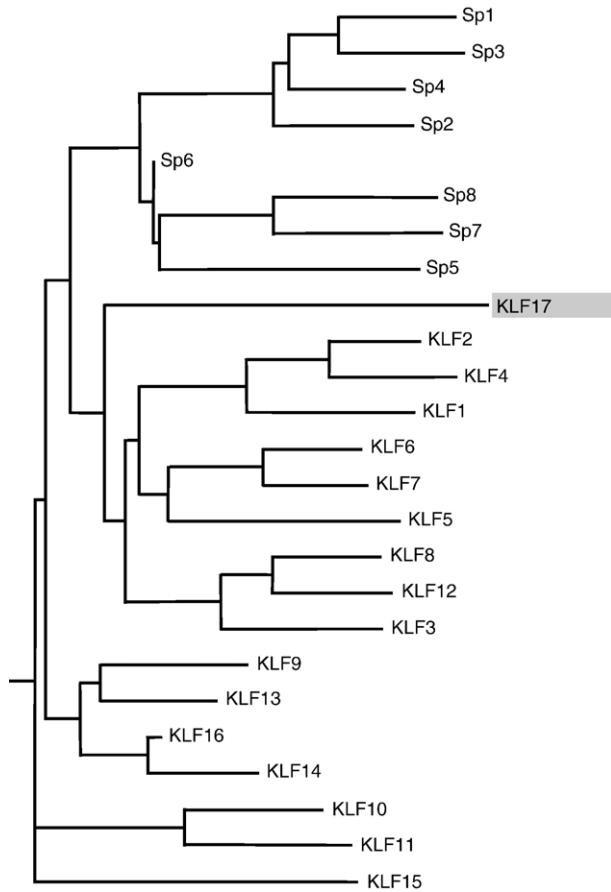
Fig. 4. The evolutionary relationship between KLF17 and the other Sp/KLFs. The full-length protein sequences of the human Sp/KLFs were used to create this phylogenetic tree.
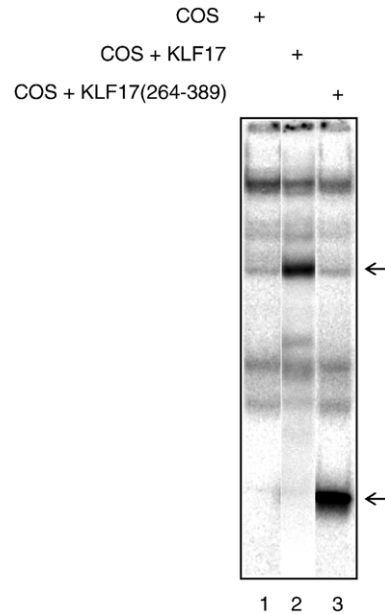

Fig. 5. KLF17 binds the β-globin CACCC box through its zinc fingers. Nuclear extracts prepared from untransfected COS cells (lane 1) or COS cells expressing full-length human KLF17 (lane 2) or the zinc fingers only (lane 3) were incubated with a β-globin promoter probe encompassing the proximal CACCC site.

report, the authors concluded that Zfp393 was not a member of either of the triple-C$_2$H$_2$ zinc finger protein families defined by the Sp/KLFs and EGR1 based only on the low sequence homology of its zinc finger domain. However, we propose that the similarities, both in sequence and in structure, between the zinc fingers of KLF17 and the Sp/KLFs, as outlined above, are sufficient to consider it a member of the family.

KLF17 is more closely related to KLF than to Sp subfamily

To clarify the relationship of human KLF17 to other members of the Sp/KLF family, the Phylogeny Analysis tool at BioManager (http://www.angis.org.au) was used to construct

an evolutionary tree of the Sp/KLFs (Fig. 4). In this analysis, KLF17 shares a common ancestor with KLFs 1–8 and 12. The Sp proteins and the remaining KLFs are on other branches of the tree.

The binding specificity of KLF17 matches that of the KLFs

The sequence similarity between the zinc fingers of KLF17 and those of other Sp/KLF members suggests that KLF17 recognizes the same DNA sequences. To verify that KLF17 can indeed bind CACCC-box sequences, electrophoretic mobility shift assays (EMSAs) were performed using overexpressed full-length human KLF17, as well as the zinc finger region alone. As predicted, KLF17 can bind a typical CACCC box (from the β-globin promoter) and the binding is mediated by the zinc fingers (Fig. 5).

The deviation of the third zinc finger of KLF17 from the consensus for the Sp/KLF family suggests that it may exhibit an altered binding specificity, or affinity. To investigate the DNA binding of KLF17 further, competition EMSA was employed

Fig. 6. DNA-binding specificity of human KLF17. (A) Full-length murine KLF3 and the zinc fingers of human KLF17 (residues 264–389) were separately expressed in COS cells. COS nuclear extracts expressing KLF17 and KLF3 were mixed. Combined extracts were incubated with increasing concentrations of an unlabeled DNA probe prior to incubation with radiolabeled β-globin CACCC-box probe. While the radiolabeled probe for all EMSAs was the β-globin CACCC-box probe, various unlabeled competitor probes were used. Unlabeled competitor probes included CACCC-box- and GC-box-containing sequences and point-mutated β-globin promoter CACCC-box sequences, in which each of the 9 bases of the CACCC box was individually mutated, and are indicated above each gel. Two portions of the same gel have been shown in each case. Note that the Sp3 band is an endogenous protein in COS cells and that the identity of this band was confirmed by Sp3 antibody supershift (results not shown). The top of the gel showing Sp1 and the larger isoform of Sp3 has been cut off. (B) The ability of KLF3, Sp3, and KLF17 to bind to the cold probes was scored. A band that became fainter with addition of increasing amounts of unlabeled competitor probe represented a band that was being bound by this probe. Hence, in this case, the binding to cold probe was scored with a plus sign. A band that did not fade with the addition of increasing amounts of unlabeled competitor probe represented a band that was not being bound by the cold probe and was scored with a minus sign. The position number of the modified nucleotides in the β-globin CACCC box has been displayed. The predicted consensus binding site for the Sp/KLF family has been included for comparison [5,29].

A
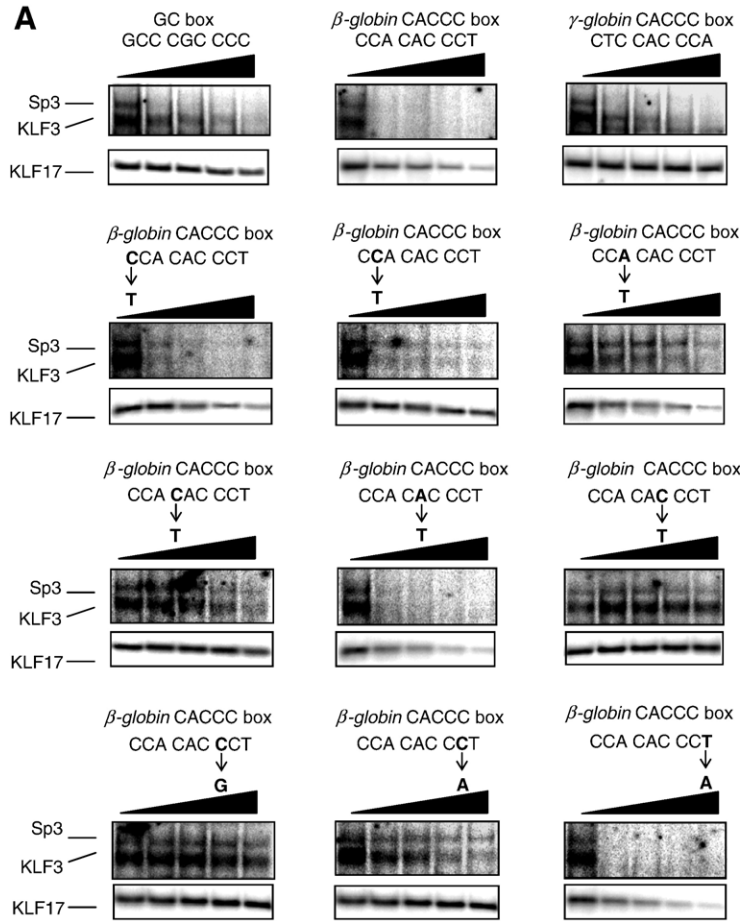
B

PROTEIN	BINDING TO COLD COMPETITOR PROBE		
	GC box	*β-globin* promoter CACCC box	*γ-globin* promoter CACCC box
Sp3	+	+	+
KLF3	+	+	+
KLF17	-	+	-

PROTEIN	BINDING TO COLD *β-GLOBIN* CACCC BOX									
	Wild type	Point mutants								
		C↓T	C↓T	A↓T	C↓T	A↓T	C↓T	C↓G	C↓A	T↓A
	Position #	1	2	3	4	5	6	7	8	9
Sp3	+	+	+	+	+	+	-	-	+	+
KLF3	+	+	+	+	+	+	-	-	+	+
KLF17	+	+	-	+	-	+	-	-	-	+
Consensus Sp/KLF site		C	C	N	C	N	C	C	C	N

Table 2
DNA-binding sites recognized by KLF17

Consensus:	N	C	C	N	C	C/G/a	C	C	C	C/g/a	N
Position:		1	2	3	4	5	6	7	8	9	
G	14	7	10	17	14	16	9	0	3	14	8
A	15	4	12	10	10	11	11	0	0	11	14
T	16	5	0	14	0	1	0	2	1	4	17
C	5	34	28	9	26	22	30	48	46	21	11

Fifty of the 64 sequences of DNA selected after three rounds of PCR site selection using GST–KLF17 fusion protein were aligned (for alignment see Supplementary Fig. 1). The number of times that each base was encountered at each position in the alignment is shown. The consensus sequence derived from this alignment is displayed at the top. The position number of each nucleotide with reference to the numbering system used for the β-globin CACCC box in Fig. 6B is displayed.

(Fig. 6A). The DNA-binding specificity of human KLF17 was compared to that of the known CACCC-box binding protein, Klf3, and the related GC-box binding protein, Sp3. Disappearance of a band with addition of increasing amounts of unlabeled competitor probe demonstrated that the protein in question was able to bind to this unlabeled probe. Where this occurred, the unlabeled probe was able to out-compete the labeled β-globin probe in binding to the protein, hence leading to a decrease in band intensity. The affinity of each protein for each competitor probe was scored and the results summarized (Fig. 6B). The KLF17 protein was able to bind appreciably only to the β-globin probe, whereas both Sp3 and Klf3 bound to the GC box and both β- and γ-globin probes. (Note: To allow the comparison of multiple panels the top of the gel has been cut off so that only the smaller isoform of Sp3 is visible.) It was particularly surprising that KLF17 was unable to bind the variant CACCC box of the γ-globin promoter as this sequence is highly related to the β-globin promoter CACCC box. Mutation of the individual bases of the β-globin CACCC box suggests that the preferred DNA-binding site of human KLF17 is approximated by the predicted consensus site for KLF1: 5'-NCNCNCCCN-3' [5,29].

To define the consensus DNA binding sequence for KLF17 better, a site selection experiment was performed. Randomly generated DNA oligonucleotides were screened for sequences that interacted with a GST–KLF17 zinc fingers fusion protein through multiple rounds of selection. Selected oligonucleotides were cloned, sequenced, and aligned (Table 2). The consensus generated was CCNC(C/G/a)CCC(C/g/a). Nucleotides flanking this 9-base consensus did not show any base preference. This consensus is an excellent fit with the CCNCNCCCN sequence, which is bound by Sp/KLF family members, further supporting the inclusion of KLF17 as a member of this family.

The consensus generated from the site selection experiment supports the EMSA results from Fig. 6. For example, mutation of the C residues at positions 2, 4, and 6 of the β-globin CACCC probe to T residues significantly reduced the ability of the probe to compete KLF17 away from labeled β-globin probe. Similarly, no T nucleotides were identified at position 2, 4, or 6 of the site selection consensus (Table 2). Note that the γ-globin

CACCC box diverges from this consensus at the base 2 (T instead of C) and may explain the lack of detectable binding by KLF17 to this sequence.

Although KLF17 binds to the β-globin CACCC probe, this binding is of a lower affinity than Klf3 binding to this probe. The site selection experiment found a low frequency of the T base at position 9. It is possible that alteration of the position 9 T to a C or G residue may increase the affinity of KLF17 for the β-globin probe and that this modified sequence might better reflect the preferred DNA-binding site of KLF17.

KLF17 activates transcription from CACCC boxes

Members of the KLF family can act as activators or repressors of transcription depending on cell and promoter context, although the molecular mechanism underlying these distinct activities remains unclear. We assessed the ability of human KLF17 to regulate transcription from β-globin CACCC-box sites using transient transfection reporter gene assays (Fig. 7). This experiment was performed in the *Drosophila* SL2 cell line, which is reportedly devoid of Sp/KLF-like activities [30]. Consequently, in these cells the basal activity of promoters under the control of G/C-rich sites is very low, and activation by transiently transfected Sp/KLFs can be seen more clearly than in other cell lines. In SL2 cells, KLF17 activated transcription of a reporter gene in a dose-dependent manner (Fig. 7). Its transactivation ability was roughly 40% of that of Klf1.

In conclusion, this report shows the DNA-binding characteristics and transactivation ability of KLF17, the human orthologue of murine Zfp393. We have demonstrated that KLF17 should be regarded as a member of the Sp/KLF family of transcription factors, based on the sequence similarity of its three C-terminal zinc fingers to those of other Sp/KLFs. In addition, phylogenetic analysis suggests it is most like the KLFs, and as such we have shown that it can regulate gene transcription from CACCC-box elements.
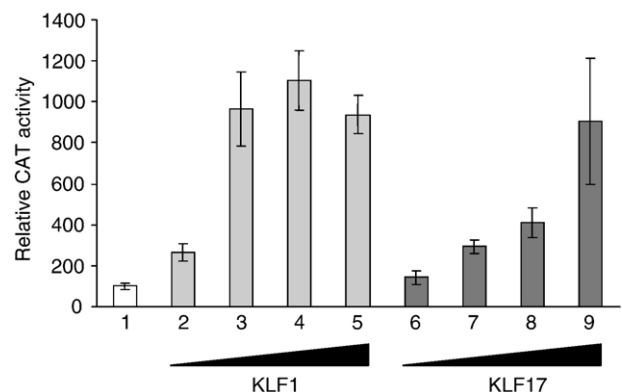

Fig. 7. Human KLF17 can activate transcription from CACCC boxes. KLF17 activates transcription of the reporter construct (CACCC-GRE)₃-Adh-CAT in SL2 cells. CAT activities are shown relative to the value for reporter only (column 1), which was set to 100. The activator KLF1 is included for comparison. The values shown are the averages of six replicates. The error bars indicate 2 standard deviations.

Materials and methods

Bioinformatics

GenBank (NCBI; http://www.ncbi.nlm.nih.gov) EST database homology searches and genomic database searches were conducted using BLAST. Evidence Viewer and Map Viewer tools were utilized for establishing the genomic structure of KLF17 and identifying murine and human orthologues.

The bioinformatics tools provided by the Australian National Genome Information Service (http://www.angis.org.au) were used for calculating percentage similarity between proteins (BestFit), sequence alignments (PileUp and BoxShade), and phylogenetic analysis (Phylogeny Analysis).

PCR and subcloning

Each coding exon of KLF17 was amplified from human genomic DNA using intronic primers to prevent amplification of ψKLF17. The exons were assembled to form a sequence that is identical to the cDNA that would be produced by the human KLF17 gene (allowing expression of the KLF17 protein). The human KLF17 gene consists of three exons. DNA encompassing these exons was amplified from human genomic DNA as follows. Exon 1 was amplified using 5′ and 3′ primers A1114, AGTAGAGAGTCTAGACCCCACC, and A1117, CG-GCCTGCCAGGGGCTGGCA; exon 2 using A1118, GAGCAAAAATCACCT-GACTCTTT, and A1121, CCTCCATCCCCACCCCACCT; and exon 3 using A1122, AAATTCCATCTCTCCCTTGTCAT, and A1124, AAAAGAAAACC-CAGAATCAGGAG. These primers correspond to intronic sequences and gene flanking regions and thereby prevent the amplification of a closely related processed pseudogene sequence. The precise exons were then amplified subsequently amplified using the material amplified above as template and using internal primers described below. The primers were overlapping so that the sequences could be joined by overlap PCR as follows. Exon 1 was amplified with a 5′ primer incorporating an upstream BamHI site, A1112, CGGGATCCATG-TACGGCCGACCGCAGGC, and a 3′ primer that overlapped the 5′ end of exon 2, A1116, TGAGTTCTCGTTATCCTGGGCAGCCTGGTGCGCC. Exon 2 was amplified using a 5′ primer that overlapped the 3′ end of exon 1, A1119, CCAGGCTGCCCAGGATAACGAGAACTCAGCGCCCAA, and a 3′ primer that overlapped the 5′ end of exon 3, A1120, AAGAATATGGCCTCT-CACCTGTGTGCTTGCGCTG. Exon 3 was amplified using a 5′ primer that overlapped the 3′ end of exon 2, A1123, AAGCACACAGGTGAGAGGCCA-TATTCTTGCAACTGG, and a 3′ primer incorporating an EcoRI site, A1113, GGAATTCCTAAGGACCAGCAGCAGGAGG. Finally after overlap PCR the full-length clone was amplified by 5′ A1112 and 3′ A1113 and inserted between the BamHI and EcoRI sites of pcDNA3Flag (Invitrogen). To construct an expression vector for Drosophila SL2 cells the cDNA was reamplified with A1112 and another primer similar to A1113 but containing a BamHI instead of an EcoRI site and inserted into the BamHI site of pPac. The cDNA is driven from the Drosophila actin 5C promoter. The zinc finger region alone (residues 264–389) was also subcloned into pcDNA3 with an N-terminal Flag tag (Invitrogen). For overexpression in Escherichia coli, amino acids 100–260 of KLF17 was subcloned into pGEX2T (Amersham Biosciences).

Cell lines

COS cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, penicillin, streptomycin, and glutamine. Schneider Line 2 (SL2) cells were maintained in Schneider's Drosophila medium supplemented with 10% heat-inactivated fetal bovine serum, penicillin, streptomycin, and glutamine.

Nuclear extract preparation and EMSA

COS cells were separately transiently transfected (FuGENE 6; Roche Molecular Biochemicals) with mammalian expression vectors for KLF17 and KLF3 and nuclear extracts were prepared [31]. The nuclear extracts from the two transfections were combined prior to addition of probes. The GC-box, γ-globin CACCC-box, and β-globin CACCC-box oligonucleotides, described previously [32], were end-labeled with [^{32}P]ATP using T4 polynucleotide kinase (New England Biolabs), and DNA binding assays performed as described [33]. For cold competition assays, increasing amounts of annealed competitor oligonucleotides (0, 40, 80, 160, and 320 ng) were added to the binding reactions and incubated at 4°C for 10 min prior to addition of ^{32}P-labeled probe (0.4 ng) and further 10-min incubation. Gels were visualized using a PhosphorImager (Molecular Dynamics) and analysis was performed using ImageQuant software (Molecular Dynamics).

Binding site selection

The method used was based on the published site selection experiment that allowed identification of the DNA binding sequence for Pegasus [34]. Control GST protein and the KLF17 DNA binding domain fusion protein, comprising amino acids 264–389 of KLF17 downstream of GST, were produced in the E. coli strain BL21. For site selection experiments a degenerate oligonucleotide with a core of 14 random nucleotides flanked by conserved sequences containing BamHI and EcoRI restriction sites was used: CAGGTCAGTT-CAGCGGATCCTGTAT-(N)$_{14}$-TATGCGAATTCAGTGCAACTGCAGC. The PCR primers used were primer F, CAGGTCAGTTCAGCGGATCCTGTAT, and primer R, GCTGCAGTTGCACTGAATTCGCATA. The random oligonucleotide was rendered double-stranded by annealing with primer R and extension with Klenow enzyme for 30 min at room temperature. Briefly, 110 μg of polyacrylamide gel purified template was mixed with 75 μg of primer R in 1× Klenow reaction buffer (50 mM Tris–HCl, pH 8.0, 10 mM MgCl$_2$, 50 mM NaCl) and annealed by boiling and cooling. Deoxynucleotides were added to a final concentration of 400 μM plus 25 μCi of [alpha-^{32}P]dCTP and 25 units of Klenow, and the mixture was incubated at room temperature for 30 min in 1× Klenow buffer. The double-stranded template was purified by polyacrylamide gel electrophoresis.

Five micrograms of GST-KLF17 (264–389) bound to GSH beads (5 μl) was incubated on ice for 30 min with 1 ng of the purified random oligonucleotide library and 10 μl of binding buffer (0.2 μg/ml of poly(dI–dC) (Sigma), 0.2 mg/ml bovine serum albumin (New England Biolabs), 25 mM Hepes, pH 7.5, 100 mM KCl, 0.1 mM ZnSO$_4$, 10 mM MgCl$_2$, 0.1% Nonidet P-40, 1 mM dithiothreitol, 5% glycerol). The beads were centrifuged and washed four times with binding buffer and boiled for 5 min in 15 μl H$_2$O. The DNA was eluted by boiling, and following 40 cycles of PCR amplification using primers F and R, the product was purified by polyacrylamide gel electrophoresis and used for further binding reactions. After three rounds of selection a gel retardation experiment was carried out using soluble fusion protein. The shifted band was excised, and the DNA was amplified as before, digested with EcoRI and BamHI, subcloned into pBluescript KS, and sequenced.

Transient transfection assays

Cells were transfected using the calcium phosphate precipitation method. SL2 cells were transfected with 500 ng of the (CACCC-GRE)$_3$-Adh-CAT reporter [35] and increasing amounts of each KLF (50, 250, 500, and 1000 ng). CAT activity was measured using the organic extraction method [36].

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ygeno.2005.12.011.

References

[1] W.S. Dynan, R. Tjian, Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II, Cell 32 (1983) 669–680.

[2] I.J. Miller, J.J. Bieker, A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins, Mol. Cell Biol. 13 (1993) 2776–2786.

[3] D.T. Dang, J. Pevsner, V.W. Yang, The biology of the mammalian Kruppel-like family of transcription factors, Int. J. Biochem. Cell Biol. 32 (2000) 1103–1121.

[4] J.J. Bieker, Kruppel-like factors: three fingers in many pies, J. Biol. Chem. 276 (2001) 34355–34358.

[5] J. Turner, M. Crossley, Mammalian Kruppel-like transcription factors: more than just a pretty finger, Trends Biochem. Sci. 24 (1999) 236–240.

[6] J. Kaczynski, T. Cook, R. Urrutia, Sp1-and Kruppel-like transcription factors, Genome Biol. 4 (2003) 206.

[7] S.A.N. Wolfe, L. Pabo, C.O. Pabo, DNA recognition by Cys2His2 zinc finger proteins, Annu. Rev. Biophys. Biomol. Struct. 3 (1999) 183–212.

[8] C.O. Pabo, E. Peisach, R.A. Grant, Design and selection of novel Cys2His2 zinc finger proteins, Annu. Rev. Biochem. 70 (2001) 313–340.

[9] W.C. Feng, C.M. Southwood, J.J. Bieker, Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor, J. Biol. Chem. 269 (1994) 1493–1500.

[10] S. Dovat, et al., A common mechanism for mitotic inactivation of C2H2 zinc finger DNA-binding domains, Genes Dev. 16 (2002) 2985–2990.

[11] D. Jantz, J.M. Berg, Reduction in DNA-binding affinity of Cys2His2 zinc finger proteins by linker phosphorylation, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 7589–7593.

[12] J.M. Shields, V.W. Yang, Two potent nuclear localization signals in the gut-enriched Kruppel-like factor define a subfamily of closely related Kruppel proteins, J. Biol. Chem. 272 (1997) 18504–18507.

[13] K. Pandya, T.M. Townes, Basic residues within the Kruppel zinc finger DNA binding domains are the critical nuclear localization determinants of EKLF/KLF-1, J. Biol. Chem. 277 (2002) 16304–16312.

[14] A. Song, et al., Functional domains and DNA-binding sequences of RFLAT-1/KLF13, a Kruppel-like transcription factor of activated T lymphocytes, J. Biol. Chem. 277 (2002) 30055–30065.

[15] W. Zhang, J.M. Shields, K. Sogawa, Y. Fujii-Kuriyama, V.W. Yang, The gut-enriched Kruppel-like factor suppresses the activity of the CYP1A1 promoter in an Sp1-dependent fashion, J. Biol. Chem. 273 (1998) 17917–17925.

[16] S. Kadam, et al., Functional selectivity of recombinant mammalian SWI/SNF subunits, Genes Dev. 14 (2000) 2441–2451.

[17] L.M. Botella, et al., Transcriptional activation of endoglin and transforming growth factor-beta signaling components by cooperative interaction between Sp1 and KLF6: their potential role in the response to vascular injury, Blood 100 (2002) 4001–4010.

[18] S. Miyamoto, et al., Positive and negative regulation of the cardiovascular transcription factor KLF5 by p300 and the oncogenic regulator SET through interaction and acetylation on the DNA-binding domain, Mol. Cell Biol. 23 (2003) 8528–8541.

[19] W. Yan, K.H. Burns, L. Ma, M.M. Matzuk, Identification of Zfp393, a germ cell-specific gene encoding a novel zinc finger protein, Mech. Dev. 118 (2002) 233–239.

[20] R.L. Strausberg, et al., Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 16899–16903.

[21] K.E. Baker, R. Parker, Nonsense-mediated mRNA decay: terminating erroneous gene expression, Curr. Opin. Cell Biol. 16 (2004) 293–299.

[22] A.J. Mighell, N.R. Smith, P.A. Robinson, A.F. Markham, Vertebrate pseudogenes, FEBS Lett. 468 (2000) 109–114.

[23] R.H. Waterston, et al., Initial sequencing and comparative analysis of the mouse genome, Nature 420 (2002) 520–562.

[24] P. Basu, et al., Evolutionary conservation of KLF transcription factors and functional conservation of human gamma-globin gene regulation in chicken, Genomics 84 (2004) 311–319.

[25] Z. Wang, et al., Gene structure and evolution of Tieg3, a new member of the Tieg family of proteins, Gene 325 (2004) 25–34.

[26] K.D. Wagner, N. Wagner, A. Schedl, The complex life of WT1, J. Cell Sci. 116 (2003) 1653–1658.

[27] Y. Shi, J.S. Lee, K.M. Galvin, Everything you have ever wanted to know about Yin Yang 1, Biochim. Biophys. Acta 1332 (1997) F49–F66.

[28] G. Thiel, G. Cibelli, Regulation of life and death by the zinc finger transcription factor Egr-1, J. Cell Physiol. 193 (2002) 287–292.

[29] R.E. Klevit, Recognition of DNA by Cys2, His2 zinc fingers, Science 253 (1991) 1367–1393.

[30] A.J. Courey, R. Tjian, Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif, Cell 55 (1988) 887–898.

[31] N.C. Andrews, D.V. Faller, A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells, Nucleic Acids Res. 19 (1991) 2499.

[32] J. van Vliet, J. Turner, M. Crossley, Human Kruppel-like factor 8: a CACCC-box binding protein that associates with CtBP and represses transcription, Nucleic Acids Res. 28 (2000) 1955–1962.

[33] M. Crossley, et al., Isolation and characterization of the cDNA encoding BKLF/TEF-2, a major CACCC-box-binding protein in erythroid cells and selected other cells, Mol. Cell Biol. 16 (1996) 1695–1705.

[34] J. Perdomo, M. Holmes, B. Chong, M. Crossley, Eos and Pegasus, two members of the Ikaros family of proteins with distinct DNA binding activities, J. Biol. Chem. 275 (2000) 38347–38354.

[35] J. Turner, M. Crossley, Cloning and characterization of mCtBP2, a co-repressor that associates with basic Kruppel-like factor and other mammalian transcriptional regulators, EMBO J. 17 (1998) 5129–5140.

[36] M.J. Sleigh, A nonchromatographic assay for expression of the chloramphenicol acetyltransferase gene in eucaryotic cells, Anal. Biochem. 156 (1986) 251–256.